

Polymorphic Admixture Typing in Human Ethnic Populations

Michael Dean,¹ J. Claiborne Stephens,¹ Cheryl Winkler,² Deborah A. Lomb,² Mark Ramsburg,² Raleigh Boaze,¹ Claudia Stewart,² Lauren Charbonneau,² David Goldman,³ Bernard J. Albaugh,⁵ James J. Goedert,⁴ R. Palmer Beasley,⁶ Lu-Yu Hwang,⁶ Susan Buchbinder,⁷ Michael Weedon,¹ Patricia A. Johnson,¹ Mary Eichelberger,¹ and Stephen J. O'Brien¹

¹Laboratory of Viral Carcinogenesis, National Cancer Institute, and ²Biological Carcinogenesis and Development Program, Program Resources, Inc./DynCorp, National Cancer Institute, Frederick Cancer Research Development Center, Frederick, MD; ³Laboratory of Clinical Studies, National Institute of Alcohol Abuse and Alcoholism, and ⁴Viral Epidemiology Branch, National Cancer Institute, National Institutes of Health, Bethesda; ⁵Institute of Behavioral Research, Clinton, OK; ⁶School of Public Health, Health Science Center, University of Texas, Houston; and ⁷San Francisco City Clinic, Department of Public Health, San Francisco

Summary

A panel of 257 RFLP loci was selected on the basis of high heterozygosity in Caucasian DNA surveys and equivalent spacing throughout the human genome. Probes from each locus were used in a Southern blot survey of allele frequency distribution for four human ethnic groups: Caucasian, African American, Asian (Chinese), and American Indian (Cheyenne). Nearly all RFLP loci were polymorphic in each group, albeit with a broad range of differing allele frequencies (δ). The distribution of frequency differences (δ values) was used for three purposes: (1) to provide estimates for genetic distance (differentiation) among these ethnic groups, (2) to revisit with a large data set the proportion of human genetic variation attributable to differentiation within ethnic groups, and (3) to identify loci with high δ values between recently admixed populations of use in mapping by admixture linkage disequilibrium (MALD). Although most markers display significant allele frequency differences between ethnic groups, the overall genetic distances between ethnic groups were small (.066–.098), and <10% of the measured overall molecular genetic diversity in these human samples can be attributed to “racial” differentiation. The median δ values for pairwise comparisons between groups fell between .15 and .20, permitting identification of highly informative RFLP loci for MALD disease association studies.

Introduction

The measurement of genetic variation in DNA markers has played a critical role in the rapid development of human genetic analysis in recent decades. DNA polymorphisms

provide the genetic markers for linkage analyses that have localized hundreds of hereditary diseases and led to the prospect of a high-resolution map of the human genome (McKusick 1991; NIH/CEPH Collaborative Mapping Group 1992, 1993; Collins and Galas 1993; Cuticchia et al. 1993). In addition, the pattern of genetic differentiation of DNA polymorphism has been applied to important questions about the natural history of human origins and has provided the foundation for the proposed human diversity project, which promises to describe the extent and character of molecular genetic variation in divergent human populations (Roberts 1992; Cavalli-Sforza et al. 1993; Gibbons 1993). Furthermore, human DNA polymorphisms are now being employed in association analysis among certain disease cohorts, in pursuit of linkage disequilibrium between DNA markers and disease loci, as a mapping approach (Chakraborty and Weiss 1988; Briscoe et al. 1994; Stephens et al. 1994 [in this issue]).

Two major categories of DNA polymorphisms have been incorporated in today's human linkage map (NIH/CEPH Collaborative Mapping Group 1992, 1993; Weissenbach et al. 1992; Murray 1993). The first type are RFLPs, which include both coding loci and anonymous DNA segments distributed throughout the human genome (Botstein et al. 1980; Bowcock et al. 1993). The second category includes short (2–5 bp) microsatellite repeats that are dispersed nearly randomly throughout the human genome and are characterized by extremely high heterozygosity levels, likely because of a rapid mutation rate by DNA replication slippage (Litt and Luty 1989; Weber and May 1989; Edwards et al. 1992; Weissenbach et al. 1992; Weber and Wong 1993). The Genome Database (Cuticchia et al. 1993) currently catalogs 7,297 human polymorphisms. Of these, 3,459 have an RFLP format, and 4,196 have a PCR format. The vast majority of these polymorphic markers have been typed only in a small panel of Caucasian-derived DNAs, with only 67 loci typed in Africans or African Americans.

In the present study, we selected 257 previously described RFLP markers distributed over the 23 human

Received February 10, 1994; accepted for publication May 31, 1994.

Address for correspondence and reprints: Dr. Stephen J. O'Brien, Chief, Laboratory of Viral Carcinogenesis, National Cancer Institute, Frederick, MD, 21702-1201.

© 1994 by The American Society of Human Genetics. All rights reserved.
0002-9297/94/5504-0022\$02.00

chromosomes (average spacing = 19.8 cM) and with reported high PIC in Caucasians. Polymorphic allele frequencies for each probe were estimated for four separate human ethnic populations—American Caucasian (CA), African American (AA), Asian (AS), and American Indian (AI)—in a sample of 1,884 chromosomes. Genetic distances between the ethnic groups were calculated, and the distribution of allele frequency difference, δ , between ethnic groups was computed. The latter value is critical for designing marker association analyses of disease cohorts. In the report by Stephens et al. (1994) that accompanies this article, we demonstrate that by selecting loci with large δ values ($\delta > .3$), linkage disequilibrium association with disease loci can be tracked over large recombination intervals (<20 cM) in recently admixed ethnic groups (e.g., AAs or Hispanics).

Finally, we estimate the proportion of an ethnic group's intrinsic genetic diversity versus that which is due to differentiation among these groups. Our results affirm and extend the conclusion, drawn in earlier studies (Lewontin 1972; Nei and Roychoudhury 1982), that <10% of the human genetic diversity in our sample is attributable to racial differentiation.

Patients, Material, and Methods

Population Samples

All CA patients were part of ongoing epidemiological cohorts collected in conjunction with clinical studies of HIV-1 infection and disease (Goedert et al. 1986, 1989; Hadler et al. 1986; Ginzburg et al. 1988; Murphy et al. 1991). AA patient samples were obtained as fibroblast or lymphocyte cultures from the Camden Cell Repository and the American Type Culture Collection.

AS samples were obtained from Chinese residents of Taiwan who were initially recruited for a large cohort study of the natural history of hepatitis B virus infection during the 1970s (Beasley et al. 1981). The majority of the subjects are government workers or their first-degree relatives residing in or near Taipei. Demographic information, including date of birth, province of family origin in China, and place of birth, was collected by Chinese nurses trained in epidemiology. The samples used for this study are from unrelated individuals.

AI samples were obtained from enrolled tribal members in Clinton, OK. Tribal origin was confirmed through voter registration records.

Patient Samples and Cell Culture

AI, CA, and AS samples were obtained as fresh blood collected in acid citrate dextrose vacutainers (Becton Dickinson). Peripheral blood mononuclear cells (PBMCs) were isolated from the blood by centrifugation over Histopaque-1007 (Sigma Diagnostics), followed by three washes at 250 g for 10 min at room temperature in (1) PBS,

(2) Iscoves modified Dulbecco's medium with 25 mM HEPES, and (3) complete medium (Iscoves supplemented with 20% FCS (Hyclone). Cell lines were established from fresh or viably frozen PBMCs and were transformed to lymphoblastoid cell lines by using Epstein-Barr virus from a B95-8 marmoset cell line (Anderson and Gusella 1984). Virus was added to cells grown in T25 flasks containing a feeder layer of irradiated cells. One-liter cultures of each patient were grown, and the cells were pelleted, washed, and frozen at -70°C.

DNA Extraction

Frozen cell pellets were thawed and lysed in 50 ml of lysis buffer containing 20 mM Tris pH 8.0, 1% SDS, RNase A. The lysate was incubated for 10 min at 60°C. Proteinase K (Boehringer Mannheim) was added to 100 µg/ml, and the incubation continued for 2 h at 37°C. The lysate was extracted with phenol:chloroform:isoamyl alcohol (10:10:1) and was ethanol precipitated; the pellet was spooled and air-dried. The DNA was resuspended in 1–5 ml of 10 mM Tris, 1 mM EDTA pH 8.0 and was resuspended overnight on a shaker. DNA was diluted with water to a concentration of 100 µg/ml.

DNA Digestions and Blots

DNA samples were assembled into 1-ml minitubes and placed on a Beckman Biomek programmable pipetting workstation. Thirty-microliter (3 µg) samples of DNA were added to microtiter plates and mixed with 10 µl of a cocktail containing restriction buffer, water, and 12 U of enzyme (for *Msp*I, 2.5 mM spermidine was added, and 18 U of enzyme were used). Plates were covered with a plastic film and incubated in a 37°C or 65°C incubator for 5–6 h. Ficoll/bromophenol loading dye was added, and the samples were loaded onto 1% agarose gels (11 × 13 cm) with a 20-well comb. A marker DNA lane (lambda cut with *Hind*III) and a blank lane were employed to allow the blot to be oriented, and electrophoresis was performed at 25 V for 16 h. Gels were stained in ethidium bromide, were photographed, and were blotted on a 0.3 × 1-m sponge soaked in 0.4 N NaOH. Blots were neutralized in 0.1 × SSC, 0.5% SDS, 200 mM Tris pH 7.5 and incubated for ≥4 h in hybridization solution (10% polyethylene glycol, 7% SDS, 5 mM EDTA, 250 mM sodium phosphate, 250 µg of single-stranded salmon sperm DNA/ml).

DNA Probes and Hybridization

Random prime labeling reactions were performed using 25 ng of supercoiled plasmid DNA and 100 µCi of ³²P-dCTP (3,000 Ci/mmol; Amersham). Labeled DNA was precipitated on ice with 2 mM spermidine and 50 µg of salmon sperm carrier DNA. DNA was pelleted and was dissolved in 0.44 ml of 500 mM NaCl, 25 mM NaOH and incubated at 37°C for 30 min. DNA probes were neutralized with 0.26 ml of 1 M Tris pH 7.6, and the specific

activity was determined by counting 2 μ l in a scintillation counter. DNA probes were added to 5 ml of hybridization solution/blot and were hybridized for 18 h at 65°C. Filters were rinsed two times with 2 \times SSC, 0.1% SDS and washed two times for 30 min each in 0.2 \times SSC, 0.1% SDS at 55–65°C. Autoradiography was performed for 24–72 h.

Data Analysis

Autoradiography results were recorded onto data-entry sheets, and the information was double-entered into a database. Information on the DNA probes is maintained in a DBase IV system. Alleles were counted on unrelated individuals for each probe, and allele frequencies were computed. Frequency and any updates will be submitted to the Genome Database (Cutticchia et al. 1993).

Statistical Analysis

Conventional population genetic statistics and tests of gene frequency differentiation were calculated as by Lewontin (1972) or Nei (1987), as follows. Expected heterozygosity (h) at a locus in each population was estimated from the allele frequency estimates (p_i ; $i = 1, 2, \dots, k$, for k alleles) as

$$h = 1 - \sum p_i^2 \quad (1)$$

and

$$h = - \sum p_i \log_2(p_i). \quad (2)$$

The former is used more widely and has a clear-cut biological meaning as gene diversity (e.g., see Nei 1987), whereas the latter has historically been used by Lewontin (1972). Contrasts of the variation within groups to that between groups were performed by unweighted averaging of allele frequencies at each marker, across populations, to give samplewide frequency estimates. The latter were used in equations (1) and (2) to estimate samplewide diversity (h_s) for comparison with results of earlier studies.

Statistical significance of gene frequency differences between populations ($\delta = X_1/n_1 - X_2/n_2$) was judged by the standard t -test, using the SD estimate $SD = [f(1-f)(1/n_1 + 1/n_2)]^{1/2}$. In the latter, f is the pooled sample frequency estimate $f = (X_1 + X_2)/(n_1 + n_2)$, where X_j and n_j are the allele count and sample size of populations $j = 1$ and $j = 2$, respectively. In addition, we have estimated genetic distance among the sampled races by using the BioSYS package to calculate Nei's (1972) standard genetic distance.

Results

Allele frequencies were estimated for 257 nuclear genetic markers in each of four human populations: CA (median sample size 938 chromosomes; range 86–1,190), AA (median sample size 250; range 4–412), AS (median sample size 88; range 2–118), and AI (median sample size 126;

range 8–164). The majority of markers (216) have two alleles, although 32 have three alleles and 9 markers have four alleles. All 257 markers were scored in the CA and AA samples, whereas 214 and 118 markers have been characterized on AS and AI, respectively. Table 1 shows the allele frequency of each marker plus the sample size for each ethnic group. Markers previously mapped on the NIH/CEPH Collaborative Mapping Group (1992, 1993) linkage map were used to define an order and approximate centimorgan intervals for each human chromosome. Additional markers were integrated into this map on the basis of recombination values in the CEPH families or, in the absence of CEPH data, by their cytogenetic localization relative to mapped markers. The resulting polymorphic admixture typing map (PATMAP) is shown in figure 1, and chromosome/order coordinates are included in table 1.

The 257 markers were selected from available polymorphic molecular clone probes on the basis of high heterozygosity and an attempt to produce equivalent linkage spacing. Thirty-eight of the RFLPs were double polymorphisms where different restriction enzymes resolved different polymorphisms with the same probe; these "two hit" RFLPs were placed at the same locus in the PATMAP (fig. 1). As the total length of the PATMAP genome is 4,337 cM (sex-averaged), the average space between loci is $4,337/(257-38) = 19.8$ cM, although certain regions (e.g., q arms of chromosomes 1–4) clearly are less dense than other regions (e.g., chromosome 11).

An important measure resolved by these data was the locus-by-locus level of allele frequency difference (δ) between ethnic groups. Loci with large δ values can be very useful for mapping genes in admixed populations (Chakraborty and Weiss 1988; Briscoe et al. 1994; Stephens et al. 1994). Table 2 presents pairwise δ values for each locus, for each ethnic group comparison. Additionally, we have indicated moderate ($>.10$) and large ($>.20$) δ values for the CA/AA comparison on the PATMAP (fig. 1). The δ values and corresponding sample sizes are large enough to achieve statistical significance ($P < .05$) in the majority of comparisons: (CA/AA, 176 of 257; CA/AS, 136 of 214; CA/AI, 89 of 118; AA/AS, 135 of 214; AA/AI, 78 of 118; and AS/AI, 65 of 115).

There may be an overall tendency for the multiallelic loci to be more highly differentiated than those with just two alleles. The latter have statistically significant differences ranging from 54% (AS/AI) to 74% (CA/AI) of loci. Loci with three alleles have statistically significant differences ranging from 69% (AS/AI) to 85% (CA/AI) of loci, and virtually all of the loci with four alleles have statistically significant differences in all population comparisons.

The magnitude of each δ value in each pairwise population comparison is associated with its PATMAP coordinate in figure 2. It is clear that high δ values are seen for many chromosomes, most chromosomal regions, and in all population comparisons. In some cases, high δ values appear to

Table I**Allele Frequencies of 257 Human RFLP Genetic Markers in Four Ethnic Groups**

CHROMOSOME (ORDER)	MARKER ^a	SYMBOL ^b	LOCATION	PROBE ^c	ENZYME ^d	SAMPLE SIZE (no. of chromosomes tested)				FREQUENCY ^e					
						NA ^f	SC ^f	CA	AA	AS	AI	CA	AA	AS	AI
1(1)	118G	PND	1p36	phANF1	BglII	2	4	900	268	116	152	.9600	.9813	1.0000	1.0000
1(2)	171H	JAK1	1	RLK	HindIII	2	8	1,014	270	58	138	.6519	.4111	.7241	.4130
1(3)	114U	D1S62	1p32	pTHI54	PvuII	2	2	1,046	332	1145526	.4759	.9825	...
1(4)	154T	D1S21	1p32-p22	5-34	TaqI	2	7	998	366	60	154	.6523	.7213	.7000	.2532
1(5)	151G	D1S18	1p32-p22	3-39	BglII	2	7	878	290	112	114	.8052	.9310	.8393	.9912
1(6)	150T	D1S17	1p32-p22	3-18	TaqI	2	7	360	2745639	.6971
1(7)	101G	D1S2	1p22	L1.22	BglII	2	1	1,128	310	114	152	.8200	.9000	.9474	.9934
1(8)	149G	D1S16	1p22-p21	2-32	BglII	2	7	830	216	114	110	.7096	.7778	.8860	.8727
1(9)	119H	D1S33	1p31-p21	p1,2-2	HindIII	2	5	1,072	314	116	160	.9170	.7834	.7931	.6938
1(9)	157H	D1S33	1p31-p21	p1,2-2	HindIII	2	5	408	2509191	.7280
1(11)	112B	D1S60	1p	pMHZ14	BamH1	2	2	1,134	308	1185908	.2403	.3220	...
1(12)	116T	D1S73	1p21-cen	pEFD53.2	TaqI	2	2	1,170	412	110	148	.6239	.8034	.2636	.4392
1(13)	132G	NGFB	1p13	phbetaN8C6	BglII	2	1	930	312	58	146	.7978	.9167	.6379	.8904
1(14)	111E	NRAS	1p13	pMCR3	EcoRI	2	2	1,062	286	1147222	.9231	.9825	...
1(15)	108M	D1S67	1q21	pHH106	MspI	2	2	1,134	338	116	162	.5212	.5148	.7759	.8827
1(16)	109M	APOA2	1q21-q23	ApollcDNA	MspI	2	3	952	274	110	98	.8225	.8869	.6182	.5816
1(17)	136M	D1S66	1q21-q31	pHBI40	MspI	2	2	776	306	546740	.1667	.2222	...
1(18)	104P	AT3	1q23-q25.1	pAT3c	PstI	2	1	946	314	92	106	.6564	.4236	.3478	.8679
1(18)	177P	AT3	1q23-q25.1	pAT1.2	PstI	2	9	408	687230	.3529
1(18)	1792	AT3	1q23-q25.1	AT3PCR	PCR	2	...	146	147192	.2857
1(20)	138H	REN	1q32	pHRnES1.9	HindIII	2	1	1,128	310	112	164	.6915	.6161	.7143	.9756
1(21)	120S	D1F10S1	1q23-q44	DR10	SstI	2	6	932	108	104	122	.7511	.8611	1.0000	.9918
2(1)	201G	D2S1	2p25	L2.30	BglII	2	1	1,124	310	88	90	.6174	.8387	.6364	1.0000
2(2)	219U	D2S12	2pter-p23	pHM20	PvuII	2	22	272	446250	.4091
2(3)	212T	D2S46	2p	pYNZ9.1	TaqI	2	2	90	165778	.7500
2(4)	221M	D2S48	2pter-q32	pEFD122	MspI	2	2	1,104	294	114	124	.5453	.7483	.8860	.7177
2(5)	204T	D2S6	2p23-p15	pXG-18	TaqI	2	74	1,146	386	110	146	.5297	.3938	.6818	.8973
2(6)	203E	D2S5	2p16-p15	IMR32-6	EcoRI	2	1	1,054	322	106	28	.7524	.7547	.6038	.6786
2(6)	203M	D2S5	2p16-p15	IMR32-6	MspI	2	1	978	90	96	148	.7096	.9111	.9896	1.0000
2(8)	232T	D2S62	2p13-q14	E135	TaqI	3	13	928	3242155	.2222
2(8)	232T	D2S62	2p13-q14			3	13	928	3247748	.7685
2(9)	211T	D2S43	2p12-cen	pYNZ15	TaqI	2	2	926	342	60	154	.5378	.7135	.4000	.4870
2(10)	233T	IL1A	2q12-q21	IL1alpha	TaqI	3	...	570	246	50	126	.3070	.1626	.0400	.0794
2(10)	233T	IL1A	2q12-q21			3	...	570	246	50	126	.4386	.5366	.2200	.0079
2(11)	202P	CRYGA	2q33-q35	pSG1	PstI	2	35	190	42	505368	.3333	.4600	...
2(12)	218T	D2S3	2q37	p5-2-96	TaqI	2	11	1,148	102	104	132	.8249	.6765	.6923	.6061
2(12)	218U	D2S3	2q37	p5-2-96	PvuII	3	11	1,186	388	882909	.0979	.0000	...
2(12)	218U	D2S3	2q37			3	11	1,186	388	885278	.5747	.6818	...
3(1)	302T	D3S31	3pter-p21	pMCT32.1	TaqI	2	2	1,028	304	1108784	.7993	.8545	...
3(1)	302U	D3S31	3pter-p21	pMCT32.1	PvuII	3	2	328	664512	.3939
3(1)	302U	D3S31	3pter-p21			3	2	328	664665	.4697
3(2)	313B	THR8	3p24.1-p22	pHEA2	BamH1	2	15	1,066	298	1166942	.4597	.8879	...
3(3)	304M	D3S30	3p14-p13	pYNZ86	MspI	2	2	978	96	46	142	.5225	.3750	.4783	.5634
3(4)	318T	D3S4	3pter-q21	B67	TaqI	2	64	768	260	60	108	.7266	.3615	.6667	.6481
3(5)	309P	CP	3q23-q25	pHCP-1	PstI	2	14	418	272	505431	.2794	.5200	...
3(6)	317B	D3S5	3q25-q28	DR-2	BamH1	2	6	170	2529647	.7976
4(1)	420G	D4S10	4p16.3	pTV20	BglII	2	1	854	244	82	144	.7412	.6721	.6585	.4792
4(1)	428M	D4S10	4p16.3	C5.5	MspI	2	20	1,116	294	1146246	.7449	.7719	...
4(2)	403M	D4S13	4p15-q21	A46	MspI	2	12	428	787687	.6667
4(3)	417H	D4S67	4p13-q13	pBS8.60	HindIII	2	18	1,098	212	88	150	.6066	.3208	.3295	.7333
4(4)	402P	AFP	4q11-q13	pHAF2	PstI	2	1	910	272	52	74	.9231	.9449	1.0000	.8514
4(5)	424E	MT2P1	4p11-q21	pHM6	EcoRI	2	35	836	310	56	20	.5861	.3645	.4643	.6000
4(6)	411E	PF4	4q12-q21	pF4	EcoRI	2	16	804	312	546878	.2083	.6296	...
4(7)	418H	IL8	4q13-q21	pMDNC F2-1,1.2	HindIII	2	19	1,016	272	116	146	.5098	.2684	.5517	.7397
4(8)	404M	ADH3	4q21-q23	pADH73	MspI	2	1	926	300	106	92	.5929	.6133	.9340	.5000
4(8)	405R	ADH2	4q21-q23	pADH36	RsaI	2	1	324	605370	.5167
4(11)	432U	ANX5	4q28-q32	EndonexinII	PvuII	2	21	740	64	885905	.8906	.3864	...

(continued)

Table I (continued)

CHROMOSOME (ORDER)	MARKER ^a	SYMBOL ^b	LOCATION	PROBE ^c	ENZYME ^d	NA ^e	SC ^f	SAMPLE SIZE (no. of chromosomes tested)				FREQUENCY ^g			
								CA	AA	AS	AI	CA	AA	AS	AI
5(1)	511M	DSS12	Sp15.2-p15.1	JO209E-B	<i>MspI</i>	2	1	1,054	334	98	130	.7230	.9132	.9082	.9846
5(2)	513E	DSS19	Sp14	JO44E-B	<i>EcoRI</i>	2	1	330	627182	.7581
5(2)	513M	DSS19	Sp14	JO44E-B	<i>MspI</i>	2	1	860	90	346686	.3778	.0588	...
5(3)	514E	DSS20	Sp13	JO71H-A	<i>EcoRI</i>	2	1	1,008	296	60	28	.6250	.6858	.9833	1.0000
5(4)	515M	DSS21	Sp13-p11	JO110H-C	<i>MspI</i>	2	1	1,150	304	1167330	.8882	.9483	...
5(5)	525M	DSS33	Sp13-p11	JO35H-A	<i>MspI</i>	2	1	748	82	...	94	.5882	.24395851
5(6)	529T	DSS76	Scen-q11.2	p105-599Ha	<i>TaqI</i>	3	1	858	250	22	54	.2809	.3080	.2727	.2037
5(6)	529T	DSS76	Scen-q11.2			3	1	858	250	22	54	.5361	.4520	.5000	.7778
5(7)	502B	DSS6	Sq11.2-q13.3	M4	<i>BamH1</i>	3	22	134	180	523134	.0722	.3462	...
5(7)	502B	DSS6	Sq11.2-q13.3			3	22	134	180	525746	.4056	.3462	...
5(8)	528M	DSS78	Sq11.2-q13.3	p105-798Rb	<i>MspI</i>	2	1	1,110	302	110	144	.5568	.2185	.3545	.5417
5(9)	523T	DSS71	Sq14-q21	C11P11	<i>TaqI</i>	2	23	880	390	588580	.6538	.8621	...
5(10)	516M	DSS22	Sq34-qter	J0205H-C	<i>MspI</i>	4	1	998	272	60	34	.1593	.0735	.0000	.0000
5(10)	516M	DSS22	Sq34-qter			4	1	998	272	60	34	.3106	.3493	.3667	.1471
5(11)	530E	DRD1	Sq34-q35	pDRD1	<i>EcoRI</i>	2	25	972	286	56	50	.8807	.8497	.7857	.9800
5(12)	504E	DSS1	S	pMLZ7.5	<i>EcoRI</i>	3	2	592	263851	.1923
5(12)	504E	DSS1	S			3	2	592	266132	.8077
6(2)	626M	D6S8	6p21.3	p2C5	<i>MspI</i>	2	1	990	100	328121	.9200	.8438	...
6(3)	624T	D6S10	6p21.3	pCH6	<i>TaqI</i>	3	1	1,080	328	86	156	.4130	.3140	.5465	.4551
6(3)	624T	D6S10	6p21.3			3	1	1,080	328	86	156	.5102	.6402	.2674	.4167
6(4)	643C	TNFA	6p21.3	phTNF5	<i>Ncol</i>	2	29	1,022	98	1086448	.5612	.4630	...
6(4)	6542	TNFB	6p21.3	Oligonucleotide	PCR	2		256	147617	.7857
6(5)	638P	HSPA1	6p21.3	ph2.3	<i>PstI</i>	2	1	1,066	308	78	120	.5854	.3734	.6667	.1750
6(6)	621G	C4A	6p21.3	pAT-A	<i>BglII</i>	2	1	942	92	60	148	.9140	.7065	.9833	1.0000
6(6)	646T	C4B	6p21.3	pAT-A	<i>TaqI</i>	3	1	1,058	356	108	150	.3185	.7893	.6296	.5000
6(6)	646T	C4B	6p21.3			3	1	1,058	356	108	150	.5794	.1517	.3704	.4933
6(7)	637T	HLA-DQA1	6p21.3	pDHC1	<i>TaqI</i>	2	27	1,070	310	98	70	.5888	.6387	.9592	.4857
6(8)	603B	D6S29	6p21	pHH157	<i>BamH1</i>	2	2	1,128	312	1125629	.1571	.5000	...
6(10)	605P	D6S41	6p21-cen	pEFD6	<i>PstI</i>	2	2	730	276	56	122	.7699	.9022	.7679	.4754
6(11)	616E	MYB	6q22-q23	pHM 2.6	<i>EcoRI</i>	2	1	1,134	384	110	36	.5106	.7943	.5636	.7222
6(11)	633E	MYB	6q22-q23	pHM 2.6	<i>EcoRI</i>	2	1	938	100	60	12	.5128	.8300	.5333	.5833
6(13)	627E	D6S2	6q27	p2-2	<i>EcoRI</i>	2	1	490	250	589837	.9840	.8103	...
6(13)	627U	D6S2	6q27	p2-2	<i>PvuII</i>	4	1	926	68	321469	.2059	.0313	...
6(13)	627U	D6S2	6q27			4	1	926	68	322829	.1618	.5625	...
6(13)	627U	D6S2	6q27			4	1	926	68	325659	.5882	.4063	...
6(14)	802T	D6S44	6q27	pYNZ132	<i>TaqI</i>	4	2	1,000	362	60	120	.1150	.0552	.1000	.3917
6(14)	802T	D6S44	6q27			4	2	1,000	362	60	120	.4110	.4751	.6000	.2333
6(14)	802T	D6S44	6q27			4	2	1,000	362	60	120	.4430	.4586	.3000	.3750
6(15)	801H	D6S39	6q27	THH5I	<i>HindIII</i>	3	2	858	74	64	162	.4382	.4324	.7500	.4012
6(15)	801H	D6S39	6q27			3	2	858	74	64	162	.5373	.4865	.2500	.5741
6(15)	801P	D6S39	6q27	THH5I	<i>PstI</i>	4	2	1,056	276	106	124	.0142	.3913	.0000	.0000
6(15)	801P	D6S39	6q27			4	2	1,056	276	106	124	.4366	.1377	.7736	.4597
6(15)	801P	D6S39	6q27			4	2	1,056	276	106	124	.5455	.4674	.2264	.5403
7(1)	721T	D7S373	7p15-p13	G80	<i>TaqI</i>	2	33	1,112	368	114	138	.8354	.6196	.7368	.7609
7(2)	729T	D7S135	7p21-p15	TM102L	<i>TaqI</i>	2	35	844	374	606896	.3075	.6167	...
7(3)	755M	D7S370	7p21-p15	pRMU7.4	<i>MspI</i>	2	2	938	254	60	26	.6876	.6102	.8333	.9231
7(4)	737H	D7S11	7p15-p13	phage 6	<i>HindIII</i>	2	1	812	82	90	114	.5296	.8171	.4000	.2719
7(5)	724U	TCRG	7p15-p14	TCR Gamma	<i>PvuII</i>	3	34	872	216	1042087	.4861	.1442	...
7(5)	724U	TCRG	7p15-p14			3	34	872	216	1046089	.4306	.6635	...
7(5)	779U	TCRGC1	7p15-p14	pC1BH0.8	<i>PvuII</i>	3	1	608	218	281908	.4083	.1071	...
7(5)	779U	TCRGC1	7p15-p14			3	1	608	218	286217	.5046	.7143	...
7(5)	780T	TCRGV11	7p15-p14	pV11SPRS	<i>TaqI</i>	2	1	274	725803	.2361
7(6)	756M	D7S371	7p12-q11	pTHH28	<i>MspI</i>	2	2	970	94	...	126	.7000	.91498571
7(7)	762HA	EGFR	7p12	pE7	<i>HindIII</i>	4	17	700	720129	.1250
7(7)	762HA	EGFR	7p12			4	17	700	721086	.1806
7(7)	762HA	EGFR	7p12			4	17	700	728729	.6250

(continued)

Table I (continued)

CHROMOSOME (ORDER)	MARKER ^a	SYMBOL ^b	LOCATION	PROBE ^c	ENZYME ^d	NA ^e	SC ^f	SAMPLE SIZE (no. of chromosomes tested)				FREQUENCY ^g			
								CA	AA	AS	AI	CA	AA	AS	AI
7(7)	762HB	EGFR	7p12	pE7	HindIII	2	17	708	725876	.6944
7(8)	739H	D7S448	7p11-q21	IEF24.11	HindIII	2	30	870	234	94	134	.5356	.6325	.2553	.2164
7(9)	760M	ERV3	7p12-q11	PHP1.7	MspI	2	36	1,070	230	1086299	.3696	.6019	...
7(10)	719E	PGY3	7q21	pMDR2	EcoRI	2	10	982	286	114	22	.7994	.5594	.7281	.7727
7(11)	702E	COL1A2	7q21.3-q22.1	NJ3 3.2	EcoRI	2	1	1,014	314	1126607	.3822	.4643	...
7(11)	703E	COL1A2	7q21.3-q22.1	NJ3 3.55	EcoRI	2	1	1,110	312	1146450	.3718	.3509	...
7(12)	707E	D7S18	7q31.1-q31.2	7c22	EcoRI	2	12	960	304	547760	.9474	1.0000	...
7(13)	704M	MET	7q31	pmet H	MspI	3	30	1,084	322	110	130	.4732	.1366	.5273	.4923
7(13)	704M	MET	7q31	...	MspI	3	30	1,084	322	110	130	.5009	.8447	.3182	.3538
7(13)	704T	MET	7q31	pmet H	TaqI	2	30	1,092	92	585211	.8152	.4310	...
7(13)	705T	MET	7q31	pmet D	TaqI	2	30	1,012	288	102	112	.8202	.9653	.7549	.9643
7(13)	706M	MET	7q31	pmet S	MspI	2	30	876	96	62	84	.9635	.9792	1.0000	.9881
7(13)	706T	MET	7q31	pmet S	TaqI	2	30	258	46	26512	.8913	.5000	...
7(14)	748M	D7S399	7q31	pB22E2.6	MspI	2	35	866	230	54	108	.6016	.1739	.5000	.2500
7(14)	749M	D7S399	7q31	pR13.5 E4.1	MspI	2	35	264	666932	.4697
7(15)	744E	D7S431	7q31	p35E5 3.0	EcoRI	2	30	974	318	629415	.9560	.9839	...
7(16)	709U	D7S8	7q31	p3H-3	PvuII	2	30	1,190	362	909311	.9088	.8444	...
7(17)	736M	D7S121	7q31-q32	d7p-141.1	MspI	2	35	816	260	609069	.5923	1.0000	...
7(18)	741H	D7S449	7q31	IEF49	HindIII	2	30	942	280	4	138	.8631	.8500	.7500	.9058
7(19)	732S	D7S117	7q31-q32	pB-48	SstI	2	35	470	76	526404	.6447	.5192	...
7(20)	733P	D7S125	7q31-q32	SA37	PstI	2	35	948	274	167869	.3978	1.0000	...
7(21)	766T	TCRB	7q35	V-BETA-97	TaqI	2	37	392	788469	.8333
7(21)	766U	TCRB	7q35	V-BETA-97	PvuII	2	37	400	788750	.9103
7(22)	740B	D7S19	7	pTS19	BamH1	2	18	1,016	84	1128907	.8214	.6964	...
7(22)	740U	D7S19	7	pTS19	PvuII	2	18	332	246	488916	.9553	1.0000	...
8(1)	806E	D8S7	8p23	pSW50	EcoRI	2	18	1,142	310	112	26	.8765	.9581	.6161	.6538
8(2)	812T	CA2	8q13-q22.1	H25-3.8	TaqI	2	1	420	76	26024	.7237	.5000	...
8(4)	805T	D8S15	8	pCR0878	TaqI	2	38	1,138	266	104	134	.5501	.2481	.8077	.5746
8(5)	808T	D8S9	8	CW1	TaqI	2	39	1,100	294	110	116	.8555	.6259	.5000	.6466
8(6)	809M	D8S13	8	Fr11-80	MspI	2	40	930	270	589151	.9407	1.0000	...
8(7)	609M	D8S19	8q	pHHH171	MspI	2	2	1,040	330	114	124	.7442	.6485	.9737	.8226
9(1)	914R	OVC	9p24	OVC2.2	RsaI	2	42	900	86	50	146	.6644	.1744	.2400	.8288
9(2)	903H	D9S3	9p21	DR6	HindIII	2	41	1,060	374	116	156	.7736	.8262	.9741	.9936
9(2)	903S	D9S3	9p21	DR6	SstI	2	41	370	72	527459	.8889	1.0000	...
9(4)	912T	D9S5	9q12-q13	DR47	TaqI	2	6	1,128	362	1148777	.8674	.6404	...
9(5)	901M	D9S15	9q13-q21.1	pEFD40.3	MspI	2	2	1,058	334	100	116	.6786	.8713	.7900	.8707
9(5)	907M	D9S15	9q13-q21.1	pMCT112	MspI	2	2	938	92	60	84	.6855	.8804	.8167	.8571
9(7)	925T	AK1	9q34.1	phAK1B3.25	TaqI	2	43	778	270	1167879	.4185	.8362	...
10(1)	1012G	RBP3	10q11.2	H.4IRBP	BglII	2	44	1,154	344	114	152	.8752	.9215	.8772	.9342
10(2)	1001U	D10S19	10q21.1-q22	pTB10.171	PvuII	2	2	474	254	526772	.9213	.7308	...
10(2)	1020U	D10S19	10q21.1-q22	pTB10.171	PvuII	2	2	1,118	104	886547	.9135	.7159	...
10(3)	1006M	D10S14	10q21.1-q23	pTHH54	MspI	3	2	762	60	564396	.2667	.8393	...
10(3)	1006M	D10S14	10q21.1-q23	...	MspI	3	2	762	60	565157	.5667	.0536	...
10(5)	1018G	D10S13	10q21.1-q23	pHHH105	BglII	2	2	988	330	112	146	.6984	.7242	.1964	.2397
10(6)	1021T	D10S1	10q22-q23	5-1	TaqI	2	1	1,134	278	1108166	.7842	.9182	...
11(1)	1134B	HRAS	11p15.5	PUC EJ 6-6	BamH1	4	1	1,048	270	1121059	.0778	.0804	...
11(1)	1134B	HRAS	11p15.5	...	BamH1	4	1	1,048	270	1121336	.2185	.0000	...
11(1)	1134B	HRAS	11p15.5	...	BamH1	4	1	1,048	270	1126679	.4481	.8750	...
11(2)	1124T	D11S12	11p15.5	pADJ762	TaqI	2	1	1,126	384	1168828	.9167	1.0000	...
11(3)	1110H	HBB	11p15.5	pRK29	HindIII	2	46	830	230	1087048	.6043	.3148	...
11(4)	1139E	D11S837E	11p15	p1596	EcoRI	2	49	922	312	406323	.6378	.7250	...
11(5)	1125T	CALCA	11p15.2-p15.1	pEM36	TaqI	2	48	380	766474	.6447
11(6)	1119P	D11S151	11p13	pS6H2.4	PstI	2	47	1,060	304	1128226	.8783	.7321	...
11(7)	623M	D11S288	11p12-p11.2	p3C7	MspI	2	1	1,004	78	1106384	.2436	.6727	...
11(8)	1116U	D11S149	11p12-p11.2	pTHH26	PvuII	2	2	1,178	386	1048514	.7202	.7308	...
11(9)	1105B	FGF3	11q13	SS6	BamH1	2	45	1,012	294	1106700	.6599	.6182	...
11(9)	1126P	FGF3	11q13	BB4	PstI	3	45	408	80	540662	.0000	.0370	...

(continued)

Table I (continued)

CHROMOSOME (ORDER)	MARKER ^a	SYMBOL ^b	LOCATION	PROBE ^c	ENZYME ^d	NA ^e	SC ^f	SAMPLE SIZE (no. of chromosomes tested)				FREQUENCY ^g			
								CA	AA	AS	AI	CA	AA	AS	AI
11(9)	1126P	FGF3	11q13			3	45	408	80	549093	.9750	.9630	...
11(10)	1113M	D11S146	11q12-q13.2	pHBI59	MspI	2	2	622	24	605482	.4583	.6500	...
11(11)	1132T	D11S84	11q22	p2-7-1D6 or p6-3	TaqI	2	11	878	262	1067244	.6412	.7736	...
11(12)	1136T	DRD2	11q22-q23	lambdaHD2G1	TaqI	3	25	1,046	302	602610	.5960	.0167	...
11(12)	1136T	DRD2	11q22-q23			3	25	1,046	302	606166	.3477	.5500	...
11(13)	1141M	CD3D	11q23	pPGBD9	MspI	2	50	480	254	189583	1.0000	1.0000	...
11(14)	1108P	D11S147	11q23.3-qter	pHBI18P2	PstI	2	2	1,018	342	52	84	.7633	.9298	.9231	.8690
11(15)	1115M	D11S144	11q22.3-q23.3	MCT128.1	MspI	2	2	1,004	94	58	144	.5438	.7660	.5345	.8403
11(16)	1104S	ETS1	11q23.3	pHE 5.4	SstI	2	1	840	250	56	98	.7488	.5520	.4107	.7041
11(16)	1133S	ETS1	11q23.3	pHE5.4	SstI	2	1	390	747308	.6081
12(1)	1210S	VWF	12p13.3-p13.2	pVWE6	SstI	2	52	92	4	485217	.7500	.5625	...
12(1)	1215G	VWF	12p13.3-p13.2	VWF110	BglII	2	26	758	258	18	70	.6003	.3140	.8889	1.0000
12(2)	1217T	CD4	12pter-p12	pT4B	TaqI	2	54	278	468094	.7174
12(3)	1226G	A2M	12p13.3-p12.3	pha2m1	BglII	2	1	802	280	112	144	.7082	.6464	.9107	.7014
12(3)	4095U	A2M	12p13.3-p12.3	pha2m1	PvuII	3	1	312	54	62244	.1296
12(3)	4095U	A2M	12p13.3-p12.3			3	1	312	54	67724	.8704	1.0000	...
12(4)	1202E	D12S2	12p12.2-p12.1	p12-16	EcoRI	2	1	426	2548333	.9370
12(5)	1219H	COL2A1	12q12-q13.1	pgHCOLIIA	HindIII	2	55	890	232	18	150	.6034	.5603	.5000	.8133
12(6)	1222T	D12S15	12q12-q24.1	pCMM1.2	TaqI	3	2	962	306	116	114	.1746	.0458	.0172	.0175
12(6)	1222T	D12S15	12q12-q24.1			3	2	962	306	116	114	.8191	.9542	.9655	.9825
12(7)	1223T	D12S16	12q12	pTHH14	TaqI	2	2	1,106	352	110	148	.6980	.9347	.8909	.8716
12(8)	1209M	D12S17	12q12-q24.1	pYNH15	MspI	3	2	966	232	961957	.7759	.3854	...
12(8)	1209M	D12S17	12q12-q24.1			3	2	966	232	966522	.1897	.5208	...
12(8)	1209R	D12S17	12q12-q24.1	pYNH15	RsaI	3	2	962	92	58	154	.3160	.7283	.4310	.7468
12(8)	1209R	D12S17	12q12-q24.1			3	2	962	92	58	154	.6767	.2500	.5517	.2532
12(9)	1205M	D12S14	12q12-q24.1	pEFD33.2	MspI	4	2	402	822139	.4024
12(9)	1205M	D12S14	12q12-q24.1			4	2	402	824030	.4268
12(10)	1208M	D12S6	12q14	p11-1-7	MspI	2	11'	1,020	102	1026461	.6961	.4412	...
12(11)	1216M	D12S8	12q14-q24.1	p7G11	MspI	2	1	980	274	112	16	.8367	.6642	.7143	.6875
12(11)	1216TA	D12S8	12q14-q24.1	p7G11	TaqI	2	1	928	88	114	124	.9634	.9773	1.0000	.9919
12(11)	1216TB	D12S8	12q14-q24.1	p7G11	TaqI	2	1	736	82	110	90	.6209	.8293	.7182	.8444
12(12)	1203T	D12S7	12q14-q24.1	pDL32B	TaqI	4	35	982	332	920967	.2410	.0109	...
12(12)	1203T	D12S7	12q14-q24.1			4	35	982	332	922678	.3524	.0761	...
12(13)	1213H	IGF1	12q22-q23	pISav3A203E	HindIII	2	53	958	248	106	140	.8309	.9556	.7264	.9857
13(1)	1315E	D13S36	13q12	pCR1360	EcoRI	2	38	1,066	294	1025038	.2583	.4804	...
13(2)	1319E	D13S6	13q12.3	pHU10	EcoRI	2	1	576	300	...	8	.8924	.65677500
13(2)	1319R	D13S6	13q12.3	pHU10	RsaI	2	1	110	16	29545	.7500	1.0000	...
13(3)	1302M	D13S1	13q13	p7F12	MspI	2	56	1,034	252	1125106	.7262	.2411	...
13(4)	1301G	D13S37	13q14	pTH162	BglII	3	2	826	302	322179	.1026	.4063	...
13(4)	1301G	D13S37	13q14			3	2	826	302	327627	.8278	.5625	...
13(5)	1307T	D13S51	13q	pMHZ17	TaqI	2	2	636	244	60	34	.8978	.9918	1.0000	1.0000
13(6)	1303M	D13S2	13q22	p9D11	MspI	2	56	812	70	565665	.6571	.8929	...
13(7)	1308E	D13S5	13q21.3-q32	pHUB8	EcoRI	2	1	114	128333	.6667
13(7)	1308H	D13S5	13q21.3-q32	pHUB8	HindIII	2	1	1,006	226	106	118	.8072	.7257	.8396	.5424
13(8)	1304M	D13S4	13q31	p1E8	MspI	2	56	274	546715	.7222
13(9)	1311T	D13S34	13q34	pCR1330	TaqI	2	38	950	248	1128284	.9476	1.0000	...
13(10)	1321H	D13S3	13q34	p9A7	HindIII	3	56	1,084	282	118	152	.3312	.2057	.6356	.6842
13(10)	1321H	D13S3	13q34			3	56	1,084	282	118	152	.6614	.7943	.3390	.3158
13(10)	1321M	D13S3	13q34	p9A7	MspI	2	56	240	50	106375	.7200	.5000	...
14(1)	1413T	TCRA	14q11.2	pY1.4	TaqI	2	34	886	2746174	.7810
14(2)	1411P	D14S14	14	Fr7-74	PstI	2	40	980	280	1146531	.2786	.4123	...
14(3)	1407P	D14S15	14	THH39	PstI	3	2	414	70	564444	.2000	.0000	...
14(3)	1407P	D14S15	14			3	2	414	70	564734	.8000	.5714	...
14(4)	1424H	D14S21	14q32.1-q32.32	pCMM62	HindIII	2	2	410	2286878	.2851
14(5)	1418E	IGHG4	14q32.33	p24BRH	EcoRI	2	1	382	70	546492	.7429	.6852	...

(continued)

Table I (continued)

CHROMOSOME (ORDER)	MARKER ^a	SYMBOL ^b	LOCATION	PROBE ^c	ENZYME ^d	NA ^e	SC ^f	SAMPLE SIZE (no. of chromosomes tested)				FREQUENCY ^g			
								CA	AA	AS	AI	CA	AA	AS	AI
14(5)	1426T	IGHD	14q32.33	CW101	TaqI	2	58	984	338	465254	.8077	.4783	...
14(6)	1402P	D14S16	14q32.32-q32.33	pTHH39	PstI	2	2	984	256	445793	.1992	.5909	...
14(7)	1420G	CHGA	14q32	Cga	BglII	2	57	668	34	60	164	.8967	.7059	.9500	1.0000
15(1)	1517R	D15S25	15	pTHH114	RsaI	2	2	268	526455	.5577
15(2)	1514E	D15S2	15q15-q22	pDP151	EcoRI	2	1	396	278	547348	.9460	1.0000	...
15(3)	1509H	D15S45	15q11-qter	pEFZ33	HindIII	2	2	986	160	108	144	.7840	.8875	.9815	.5625
15(4)	1508T	D15S44	15q11-qter	pEFD52.1	TaqI	2	2	270	487259	.3958
15(5)	1506B	D15S28	15	pYNZ90.1	BamH1	2	2	1,086	294	1128214	.9524	.8214	...
15(7)	1505M	D15S27	15	pTHH55	MspI	2	2	942	92	60	150	.7176	.7500	.9167	.8333
15(8)	1507E	D15S37	15	pEFD85.7	EcoRI	2	2	1,044	302	64	14	.5527	.5762	.8281	.4286
15(9)	1529T	IGF1R	15q25-qter	pIGF-I-R.8	TaqI	2	59	860	78	489779	.6667	1.0000	...
16(1)	1625G	D16S37	16pter-p13	16/02	BglII	2	1	648	24	66	106	.9537	.9167	.9394	.9811
16(1)	1625M	D16S37	16pter-p13	16/02	MspI	3	1	812	288	962734	.0625	.0000	...
16(1)	1625M	D16S37	16pter-p13			3	1	812	288	966786	.8472	1.0000	...
16(2)	1608T	D16S4	16q22.1	ACH207	TaqI	2	60	980	378	107388	.7831	.8000	...
16(3)	1602E	HP	16q22.1	hp2alpha	EcoRI	2	1	368	745326	.5000
16(3)	1602H	HP	16q22.1	hp2alpha	HindIII	2	1	1,086	340	112	160	.6087	.4971	.7232	.5125
17(1)	1718H	MYH2	17p13.1	p10-5	HindIII	2	1	962	288	114	134	.7100	.6806	.3421	.3806
17(2)	1741G	SCYA@	17	Act2	BglII	2	...	92	148	527717	.7500	.7115	...
17(3)	1742H	SCYA1	17	I309	HindIII	2	...	350	685629	.3971
17(4)	1716P	D17S80	17q	pYNZ94	PstI	2	2	1,026	280	56	108	.7515	.9286	1.0000	.9815
17(6)	1719M	PPY	17p11.1-qter	pGEM-pp	MspI	2	62	824	80	825716	.5750	.9024	...
17(7)	1721M	D17S21	17q23-qter	pC63	MspI	2	11	988	276	114	16	.6265	.5471	.8333	.5625
17(7)	1733M	D17S21	17q23-qter	pC63	MspI	2	1	1,118	308	112	110	.6315	.5357	.8214	.8818
18(1)	1816T	D18S3	18p11.3	B74	TaqI	2	64	938	292	60	52	.8145	.9486	.9167	.9038
18(2)	1802P	D18S6	18q12.3	L2.7	PstI	2	1	1,028	256	104	110	.7743	.8945	.3462	.5273
18(3)	1808P	D18S5	18q22	OS-4	PstI	2	63	1,114	108	805736	.4352	.5000	...
18(3)	1808T	D18S5	18q22	OS-4	TaqI	2	63	1,050	240	1027857	.5708	.5882	...
19(1)	1907R	INSR	19p13.3	pHIR/P12-1	RsaI	2	1	292	585479	.2414
19(2)	1912T	C3	19p13.3-p13.2	pC3	TaqI	2	68	1,096	362	56	154	.9462	.9972	.9643	.9870
19(3)	1901H	D19S11	19p13.1	p13.1-82	HindIII	2	11	1,014	270	96	158	.8511	.8111	.9896	.8481
19(4)	1903M	D19S7	19q12	p4.1	MspI	2	65	918	276	1145654	.4312	.6930	...
19(5)	1902E	D19S9	19q13.1	p1J2	EcoRI	2	65	1,024	282	56	18	.9023	.9326	1.0000	.5556
19(6)	1905S	D19S8	19q13.2	p17.1	SstI	2	65	522	76	485498	.7500	.8542	...
19(7)	1906P	D19S19	19q13.2	LDR152	PstI	3	66	998	318	108	76	.1313	.4277	.5000	.3947
19(7)	1906P	D19S19	19q13.2			3	66	998	318	108	76	.8617	.5723	.5000	.6053
19(8)	1913T	APOC2	19q13.2	CII-E	TaqI	2	69	1,042	234	116	140	.5633	.3846	.3793	.1643
19(9)	1911T	CKM	19q13.3	pJN2CK-M	TaqI	2	67	1,028	288	106	66	.7004	.7882	.1509	.0909
20(1)	2002M	D20S5	20p12	pR12.21	MspI	2	65	414	826522	.7561
20(2)	2001T	D20S6	20p12	pD3H12	TaqI	2	65	1,044	272	1045239	.4265	.1154	...
20(3)	2003B	D20S14	20p11.22-p11.21	p4.8	BamH1	4	24	1,092	350	1140201	.0029	0.0000	...
20(3)	2003B	D20S14	20p11.22-p11.21			4	24	1,092	350	1143910	.5571	0.0000	...
20(3)	2003B	D20S14	20p11.22-p11.21			4	24	1,092	350	1145833	.4200	1.0000	...
20(4)	2014M	D20S23	20p12-p11.2	pFMS76	MspI	2	70	1,084	330	96	140	.8054	.9121	.7708	.6929
20(5)	2008M	D20S4	20q13.2	pMS1-27	MspI	2	1	1,064	100	112	124	.5940	.6200	.6607	.9839
21(1)	2101G	D21S26	21q11	26c	BglII	3	1	876	54	58	124	.4315	.2407	.5000	.1774
21(1)	2101G	D21S26	21q11			3	1	876	54	58	124	.5183	.6111	.4828	.7823
21(2)	2104T	D21S11	21q21	pG95-alpha1-11a	TaqI	2	2	838	84	467112	.6667	.5217	...
21(3)	2110B	D21S82	21q22.1-qter	Fr8-77	BamH1	3	40	680	270	922191	.2259	.4022	...
21(3)	2110B	D21S82	21q22.1-qter			3	40	680	270	925824	.2778	.2283	...
21(5)	2103M	SOD1	21q22.1	pS61-10	MspI	2	1	990	296	62	130	.6859	.6014	.6935	.9769
21(6)	2105MA	D21S15	21q22.3	pGSE8	MspI	2	71	1,090	96	60	154	.6018	.7917	.6500	.8506
21(6)	2105MB	D21S15	21q22.3	pGSE8	MspI	2	71	614	20	56	136	.7117	.8500	.6429	.6985
21(7)	2106P	D21S19	21q22.3	pGSB3	PstI	3	71	344	2961250	.0270
21(7)	2106P	D21S19	21q22.3			3	71	344	2968314	.9459
22(1)	2201P	D22S10	22q11.1-q11.2	22C1-18	PstI	2	1	816	90	506360	.6222	.4800	...
22(1)	2201T	D22S10	22q11.1-q11.2	22C1-18	TaqI	2	1	1,030	240	1025718	.3917	.3824	...

(continued)

Table I (continued)

CHROMOSOME (ORDER)	MARKER ^a	SYMBOL ^b	LOCATION	PROBE ^c	ENZYME ^d	NA ^e	SC ^f	SAMPLE SIZE (no. of chromosomes tested)				FREQUENCY ^g			
								CA	AA	AS	AI	CA	AA	AS	AI
22(2)	2205H	PDGFB	22q12.3-q13.1	v-SISR12	HindIII	2	72	228	467149	.7609
22(3)	2207S	D22S15	22q12-q13	DP22	SstI	2	73	86	4	187093	.2500	.8333	...
23(1)	2304U	DXS43	Xp22.2	pD2	PvuII	2	1	958	72	886347	.9306	.9659	...
23(2)	2311P	DXS41	Xp22.1	p99-6	PstI	2	78	778	100	84	82	.5771	.5700	.7381	.8293
23(3)	2317H	DXS92	Xp22.1	pXG-16	HindIII	3	74	936	84	54	144	.4145	.2976	.0000	.5069
23(3)	2317H	DXS92	Xp22.1			3	74	936	84	54	144	.5417	.7024	.8148	.3750
23(4)	2322M	DXS14	Xp11.21	p58-1	MspI	2	1	932	100	307028	.9300	.5667	...
23(5)	2324T	DXS1	Xq11.2-q12	p8	TaqI	2	1	914	92	368315	.6196	1.0000	...
23(6)	2316T	DXYS1X	Xq21.31	pDP34	TaqI	3	1	1,060	104	1083160	.4904	.1574	...
23(6)	2316T	DXYS1X	Xq21.31			3	1	1,060	104	1084660	.3846	.2870	...
23(7)	2310P	DXYS2X	Xq21.3	7b	PstI	2	77	316	645601	.5781
23(8)	2319G	DXS87	Xq21.33-q22	pA13.RI	BglII	2	65	938	64	58	150	.7889	.5156	.6897	.9867
23(9)	2306P	DXS94	Xq22	pXG-12	PstI	2	74	1,032	96	104	146	.5058	.5625	.4712	.3288
23(10)	2303T	DXS11	Xq24-q25	p22-33	TaqI	2	1	1,024	84	114	144	.9131	.3690	.8070	.7361
23(11)	2325T	DXS10	Xq26.1	36B-2	TaqI	2	1	1,064	82	1105395	.6585	.4364	...
23(12)	2312T	DXS51	Xq26.2-q26.3	pS2A	TaqI	2	1	364	746621	.7297
23(13)	2307T	DXS102	Xq26.3-q27.1	cX38.1	TaqI	3	75	750	72	56	136	.0880	.0417	.2500	.3897
23(13)	2307T	DXS102	Xq26.3-q27.1			3	75	750	72	56	136	.8600	.9583	.7500	.6103
23(15)	2314M	DXS98	Xq27.2	4D-8	MspI	2	1	1,038	106	58	138	.8391	.8491	.8621	.9928
23(16)	2504M	DXF1S2	Xq26-qter	pAX-6	MspI	3	1	410	744220	.5270
23(16)	2504M	DXF1S2	Xq26-qter			3	1	410	745415	.4595
23(17)	2308G	DXS15	Xq28	DX13	BglII	2	76	982	90	92	152	.5713	.8000	.7283	.9474

^a Laboratory designation of locus.^b Human symbol based on Human Gene Mapping nomenclature (McAlpine et al. 1993).^c Name of molecular clone.^d Restriction enzyme used to resolve RFLP by Southern blot analysis.^e No. of alleles resolved by the probe/enzyme combination.

^f Source code for source of molecular clone: 1 = ATCC; 2 = R. White; 3 = J. Scott; 4 = P. Frossard; 5 = A. I. Grossman; 6 = A. J. Driesel; 7 = N. Dracopoli; 8 = Z. Howard; 9 = E. Prochownich; 10 = I. Roninson; 11 = M. Litt; 12 = R. Williamson; 13 = L. Warnich; 14 = R. MacGillivray; 15 = B. Vennstrom; 16 = M. Poncz; 17 = G. Merlino; 18 = S. Wood; 19 = N. Mukaida; 20 = M. Skraastad; 21 = M. Jaye; 22 = E. Dietzsch; 23 = K. Nyberg; 24 = N. K. Spurr; 25 = D. Grandy; 26 = H. Pannekoek; 27 = J. Strominger; 29 = T. Spies; 30 = M. Dean; 33 = G. Camerino; 34 = T. Mak; 35 = L. C. Tsui; 36 = M. Cohen; 37 = M. Robinson; 38 = A. E. Bale; 39 = K. Olek; 40 = G. Scherer; 41 = K. Grzeschik; 42 = D. Blair; 43 = N. Bech-Hansen; 44 = C. D. Bridges; 45 = G. Peters; 46 = R. Kaufman; 47 = V. Huff; 48 = J. Hoppener; 49 = J. Lichy; 50 = C. Terhorst; 52 = D. Ginsburg; 53 = M. Jansen; 54 = R. Axel; 55 = C. M. Strom; 56 = W. Cavenee; 57 = M. Levine; 58 = F. Blattner; 59 = A. Ullrich; 60 = V. Hyland; 62 = T. Takeuchi; 63 = I. Nishisho; 64 = J. Mandel; 65 = D. Shaw; 66 = A. D. Roses; 67 = R. Dottin; 68 = G. Fey; 69 = O. Myklebost; 70 = F.-M. Stoltz; 71 = G. D. Stewart; 72 = K. Robbins; 73 = D. Kurnit; 74 = P. Szabo; 75 = P. Pearson; 76 = K. Davies; 77 = J. Weissenbach; and 78 = L. Kunkel.

^g Frequency of most common CA allele in each ethnic group is presented. For two-allele polymorphism, one line is used; for three-allele RFLP, two lines are used; for four-allele RFLP, three lines are used.

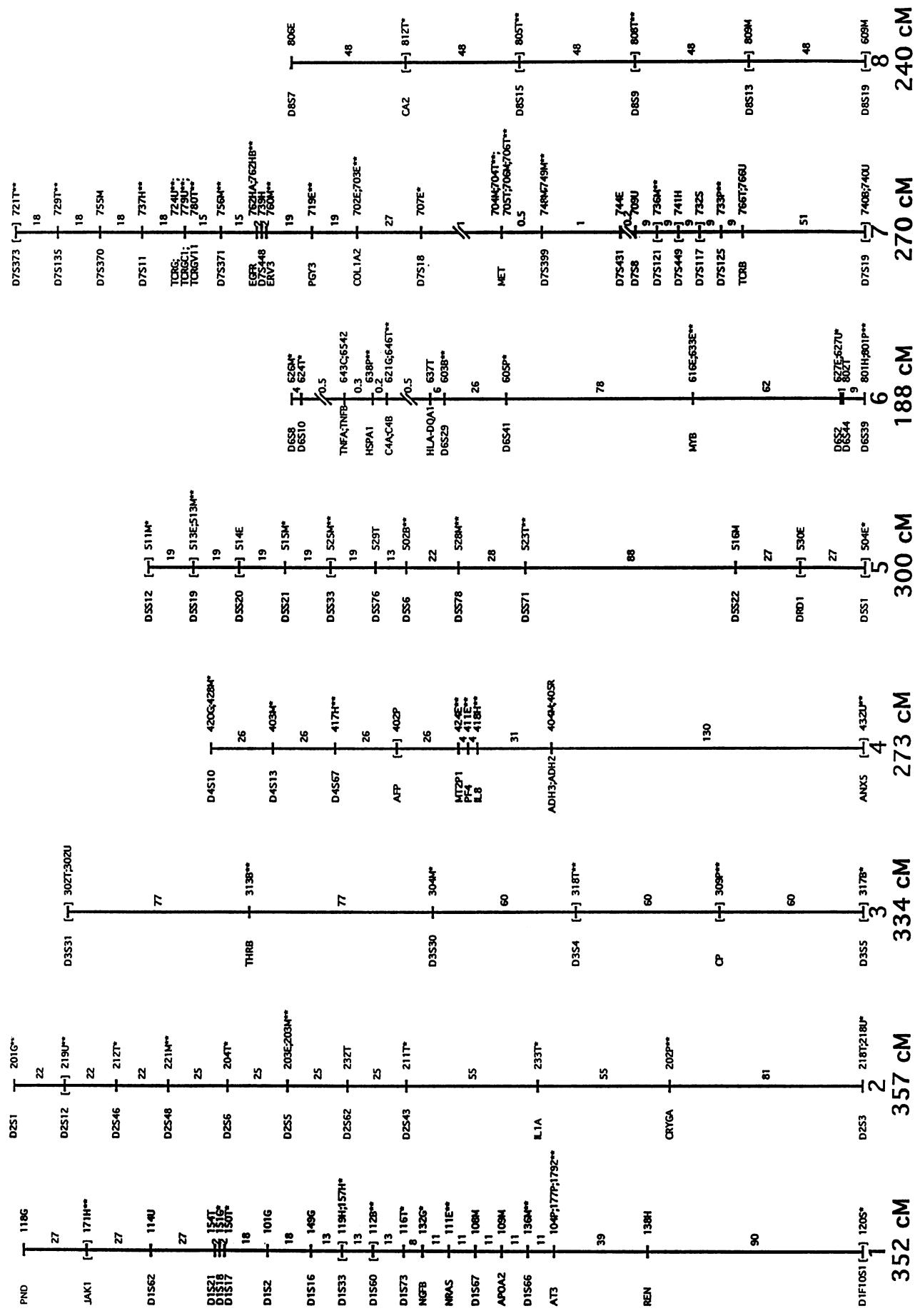
cluster together, although a formal analysis would be premature until the locus density is more uniform.

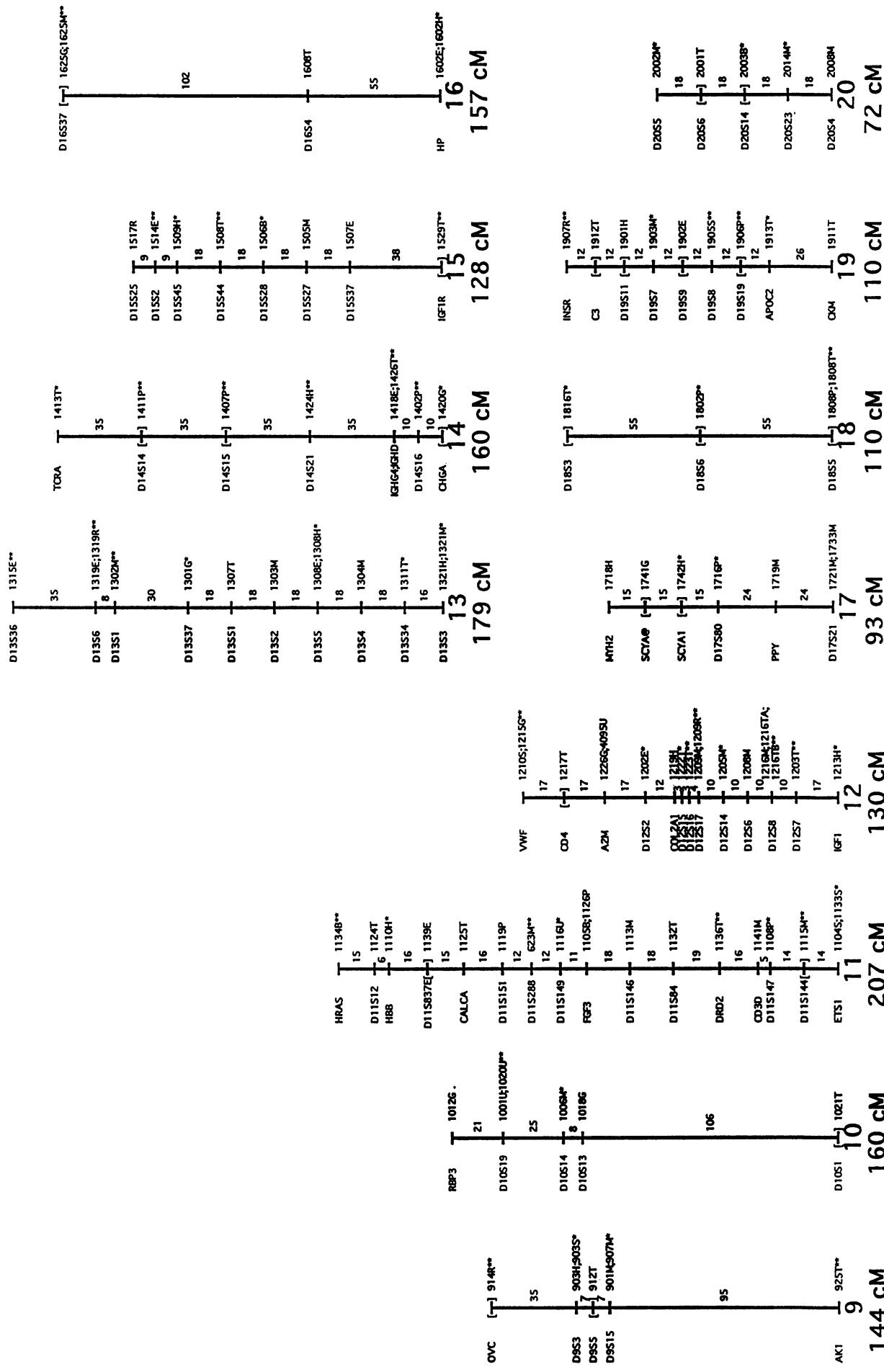
The overall level of genetic differentiation among populations is quite considerable. Figure 3 shows histograms of δ values for each population comparison. The proportion of markers having $\delta > .2$ in population comparisons CA/AA, CA/AS, and AA/AI ranges from 34% to 37%, whereas CA/AI, AA/AS, and AS/AI show 46% of markers with $\delta > .2$.

The data in table 1 can be used to contrast within-group variability to that which can be attributed to differentiation among the groups. In a study of total human genetic variation, Lewontin (1972) had estimated that a mean 85.4% of genetic variation was found within ethnic populations, leaving ~15% of the variation as reflecting differ-

entiation between ethnic groups. His estimates, using the approximation of allelic frequency heterozygosity from equation (2) were based on an analysis of 17 allozyme and blood group loci. Nei and Roychoudhury (1982) reached a similar conclusion (9%–11% of human variation is attributed to racial differentiation) on the basis of a larger data set of 85 biochemical loci.

We have estimated the average heterozygosity over all loci tested in each ethnic group (h_p) by using equations (1) and (2) (table 3). Further, we combined all the data in table 1 to estimate heterozygosity for the entire sample of four ethnic groups (h_s) on the basis of the 257 RFLP markers (table 3). The fraction of overall variation attributed to differentiation between ethnic groups equals $1 - (h_p/h_s)$ for each locus (h_p = heterozygosity averaged over all ethnic





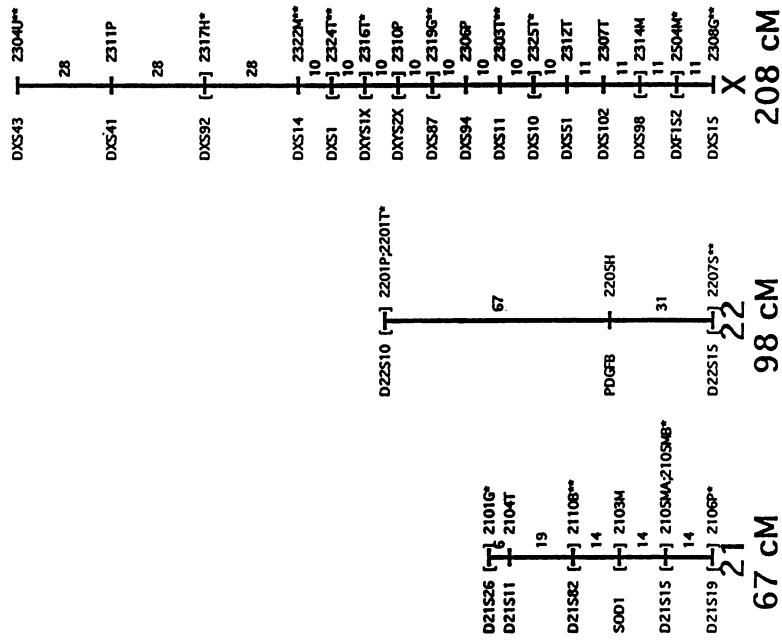


Figure 1 PATMAP. Spacing of loci is based on the NIH/CEPH Collaborative Mapping Group (1992, 1993) map and on linkage to CEPH markers, for those tested on CEPH families. Polymorphic loci not included in CEPH data, but with cytogenetic localization only, were placed in an arbitrary position on the PATMAP, relative to markers on the CEPH map, with linkage and cytogenetic locales. These loci are indicated by square brackets ([]). The Human Gene Mapping locus symbol is on the left; our laboratory designation is on the right. Multiple laboratory marker designation indicates that two different restriction enzymes give RFLPs but reside at the same locus position at this resolution level. Chromosomes are oriented with the p arm on top. * = Moderate ($\delta > .1$) differentiation between CA and AA. ** = High ($\delta > .2$) differentiation between CA and AA.

Table 2**δ Values for 257 Human RFLP Genetic Markers in Pairwise Combinations of Four Ethnic Groups**

DNA	CA/AA	CA/AS	CA/AI	AA/AS	AA/AI	AS/AI
118G0213	.0400	0.400	.0187	.0187	.0000
171H2408	.0722	.2389	.3130	.0019	.3111
114U0767	.42995066
154T0690	.0477	.3991	.0213	.4681	.4468
151G1258	.0341	.1860	.0917	.0602	.1519
150T1332
101G0800	.1274	.1734	.0474	.0934	.0460
149G0682	.1764	.1631	.1082	.0949	.0133
119H1336	.1239	.2232	.0097	.0896	.0993
157H1911
112B3505	.26880817
116T1795	.3603	.1847	.5398	.3642	.1756
132G1189	.1599	.0926	.2788	.0263	.2525
111E2009	.26030594
108M0064	.2547	.3615	.2611	.3679	.1068
109M0644	.2043	.2409	.2687	.3053	.0366
136M5073	.45180555
17924335
104P2328	.3086	.2115	.0758	.4443	.5201
177P3701
138H0754	.0228	.2841	.0982	.3595	.2613
120S1100	.2489	.2407	.1389	.1307	.0082
201G2213	.0190	.3826	.2023	.1613	.3636
219U2159
212T1722
221M2030	.3407	.1724	.1377	.0306	.1683
204T1359	.1521	.3676	.2880	.5035	.2155
203E0023	.1486	.0738	.1509	.0761	.0748
203M2015	.2800	.2904	.0785	.0889	.0104
232T0067
211T1757	.1378	.0508	.3135	.2265	.0870
233T1444	.4856	.6583	.4392	.6119	.2121
202P2035	.07681267
218T1484	.1326	.2188	.0158	.0704	.0862
218U1930	.29091071
302T0791	.02390552
302U0573
313B2345	.19374282
304M1475	.0442	.0409	.1033	.1884	.0851
318T3651	.0599	.0785	.3052	.2866	.0186
309P2637	.02312406
317B1671
420G0691	.0827	.2620	.0136	.1929	.1793
428M1203	.14730270
403M1020
417H2858	.2771	.1267	.0087	.4125	.4038
402P0218	.0769	.0717	.0551	.0935	.1486
424E2216	.1218	.0139	.0998	.2355	.1357
411E4795	.05824213
418H2414	.0419	.2299	.2833	.4713	.1880
404M0204	.3411	.0929	.3207	.1133	.4340
405R0203
432U3001	.20415042
511M1902	.1852	.2616	.0050	.0714	.0764
513E0399
513M2908	.60983190
514E0608	.3583	.3750	.2975	.3142	.0167
515M1552	.21530601
525M344300313412	...

(continued)

Table 2 (continued)

DNA	CA/AA	CA/AS	CA/AI	AA/AS	AA/AI	AS/AI
529T0841	.0443	.2417	.0480	.3258	.2778
502B4102	.22842740
528M3383	.2023	.0151	.1360	.3232	.1872
523T2042	.00412083
516M0858	.1840	.4721	.1568	.4522	.5529
530E0310	.0950	.0993	.0640	.1303	.1943
504E1945
626M1079	.03170762
624T1300	.2428	.0935	.3728	.2235	.1493
65420240
643C0836	.18180982
638P2120	.0813	.4104	.2933	.1984	.4917
621G2075	.0693	.0860	.2768	.2935	.0167
646T4708	.3111	.1815	.2187	.3416	.1296
637T0499	.3704	.1031	.3205	.1530	.4735
603B4058	.06293429
605P1323	.0020	.2945	.1343	.4268	.2925
616E2837	.0530	.2116	.2307	.0721	.1586
633E3172	.0205	.0705	.2967	.2467	.0500
627E0003	.17341737
627U1211	.27964007
802T0641	.1890	.2767	.1586	.3365	.3667
801H0566	.3118	.0370	.3176	.0876	.3488
801P3771	.3370	.0231	.6359	.3913	.3139
721T2158	.0986	.0745	.1172	.1413	.0241
729T3821	.07293092
755M0774	.1457	.2355	.2231	.3129	.0898
737H2875	.1296	.2577	.4171	.5452	.1281
724U2774	.06453419
779U2175	.09263012
780T3442
756M214915710578	...
762HA2479
762HB1068
739H0969	.2803	.3192	.3772	.4161	.0389
760M2603	.02802323
719E2400	.0713	.0267	.1687	.2133	.0446
702E2785	.19640821
703E2732	.29410209
707E1714	.22400526
704M3438	.1827	.1471	.5265	.4909	.0356
704T2941	.09013842
705T1451	.0653	.1441	.2104	.0010	.2094
706M0157	.0365	.0246	.0208	.0089	.0119
706T2401	.15123913
748M4277	.1016	.3516	.3261	.0761	.2500
749M2235
744E0145	.04240279
709U0223	.08670644
736M3146	.09314077
741H0131	.1131	.0427	.1000	.0558	.1558
732S0043	.12121255
733P3891	.21316022
766T0136
766U0353
740B0693	.19431250
740U0637	.10840447
806E0816	.2604	.2227	.3420	.3043	.0377
812T1213	.10242237
805T3020	.2576	.0245	.5596	.3265	.2331

Table 2 (continued)

DNA	CA/AA	CA/AS	CA/AI	AA/AS	AA/AI	AS/AI
808T2296	.3555	.2089	.1259	.0207	.1466
809M0256	.08490593
609M0957	.2295	.0784	.3252	.1741	.1511
914R4900	.4244	.1644	.0656	.6544	.5888
903H0526	.2005	.2200	.1479	.1674	.0195
903S1430	.25411111
912T0103	.23732270
901M1927	.1114	.1921	.0813	.0006	.0807
907M1949	.1312	.1716	.0637	.0233	.0404
925T3694	.04834177
1012G0463	.0020	.0590	.0443	.0127	.0570
1001U2441	.05361905
1020U2588	.06121976
1006M1729	.46215726
1018G0258	.5020	.4587	.5278	.4845	.0433
1021T0324	.10161340
1134B2198	.20714269
1124T0339	.11720833
1110H1005	.39002895
1139E0055	.09270872
1125T0027
1119P0557	.09051462
623M3948	.03434291
1116U1312	.12060106
1105B0101	.05180417
1126P0662	.05370370
1113M0899	.10181917
1132T0832	.04921324
1136T3350	.31095793
1141M0417	.04170000
1108P1665	.1598	.1057	.0067	.0608	.0541
1115M2222	.0093	.2965	.2315	.0743	.3058
1104S1968	.3381	.0447	.1413	.1521	.2934
1133S1227
1210S2283	.04081875
1215G2863	.2886	.3997	.5749	.6860	.1111
1217T0920
1226G0618	.2025	.0068	.2643	.0550	.2093
4095U0980	.22761296
1202E1037
1219H0431	.1034	.2099	.0603	.2530	.3133
1222T1351	.1574	.1634	.0286	.0283	.0173
1223T2367	.1929	.1736	.0438	.0631	.0193
1209M5802	.18973905
1209R4267	.1250	.4308	.3017	.0217	.3158
1205M1925
1208M0500	.20492549
1216M1725	.1224	.1492	.0501	.0233	.0268
1216TA0139	.0366	.0285	.0227	.0146	.0081
1216TB2084	.0973	.2235	.1111	.0151	.1262
1203T2369	.19173535
1213H1247	.1045	.1548	.2292	.0301	.2593
1315E2453	.02342219
1319E235714240933	...
1319R2045	.04552500
1302M2156	.26954851
1301G1153	.20023037
1307T0940	.1022	.1022	.0082	.0082	.0000
1303M0906	.32642358

(continued)

Table 2 (continued)

DNA	CA/AA	CA/AS	CA/AI	AA/AS	AA/AI	AS/AI
1308E1666
1308H0815	.0324	.2648	.1139	.1833	.2972
1304M0507
1311T1192	.17160524
1321H1329	.3224	.3530	.4553	.4785	.0486
1321M0825	.13752200
1413T1636
1411P3745	.24081337
1407P3266	.44444286
1424H4027
1418E0937	.03600577
1426T2823	.04713294
1402P3801	.01163917
1420G1908	.0533	.1033	.2441	.2941	.0500
1517R0878
1514E2112	.26520540
1509H1035	.1975	.2215	.0940	.3250	.4190
1508T3301
1506B1310	.00001310
1505M0324	.1991	.1157	.1667	.0833	.0834
1507E0235	.2754	.1241	.2519	.1476	.3995
1529T3112	.02213333
1625G0370	.0143	.0274	.0227	.0644	.0417
1625M2109	.32141528
1608T0443	.06120169
1602E0326
1602H1116	.1145	.0962	.2261	.0154	.2107
1718H0294	.3679	.3294	.3385	.3000	.0385
1741G0217	.06020385
1742H1658
1716P1771	.2485	.2300	.0714	.0529	.0185
1719M0034	.33083274
1721M0794	.2068	.0640	.2862	.0154	.2708
1733M0958	.1899	.2503	.2857	.3461	.0604
1816T1341	.1022	.0893	.0319	.0448	.0129
1802P1202	.4281	.2470	.5483	.3672	.1811
1808P1384	.07360648
1808T2149	.19750174
1907R3065
1912T0510	.0181	.0408	.0329	.0102	.0227
1901H0400	.1385	.0030	.1785	.0370	.1415
1903M1342	.12762618
1902E0303	.0977	.3467	.0674	.3770	.4444
1905S2002	.30441042
1906P2964	.3687	.2634	.0723	.0330	.1053
1913T1787	.1840	.3990	.0053	.2203	.2150
1911T0878	.5495	.6095	.6373	.6973	.0600
2002M1039
2001T0974	.40853111
2003B1661	.41675800
2014M1067	.0346	.1125	.1413	.2192	.0779
2008M0260	.0667	.3899	.0407	.3639	.3232
2101G1908	.0685	.2640	.2593	.1712	.3226
2104T0445	.18951450
2110B3046	.35411763
2103M0845	.0076	.2910	.0921	.3755	.2834
2105MA1899	.0482	.2488	.1417	.0589	.2006
2105MB1383	.0688	.0132	.2071	.1515	.0556
2106P1145

(continued)

Table 2 (continued)

DNA	CA/AA	CA/AS	CA/AI	AA/AS	AA/AI	AS/AI
2201P0138	.15601422
2201T1801	.18940093
2205H0460
2207S4593	.12405833
2304U2959	.33120353
2311P0071	.1610	.2522	.1681	.2593	.0912
2317H1607	.4145	.1667	.2976	.3274	.5069
2322M2272	.13613633
2324T2119	.16853804
2316T1744	.33764306
2310P0180
2319G2733	.0992	.1978	.1741	.4711	.2970
2306P0567	.0346	.1770	.0913	.2337	.1424
2303T5441	.1061	.1770	.4380	.3671	.0709
2325T1190	.10312221
2312T0676
2307T0983	.1620	.3017	.2083	.3480	.1397
2314M0100	.0230	.1537	.0130	.1437	.1307
2504M1050
2308G2287	.1570	.3761	.0717	.1474	.2191

NOTE.— δ Values represent the largest allele frequency difference at the marker for each population comparison. For example, for marker 118G, the entries of CA/AA are $|.9600 - .9813|$ (from the first row of table 1) = .0213. For the same marker, the entries of CA/AS are $|.9600 - 1.000| = .0400$, etc.

populations in the comparison). The average of this quantity over all loci typed provides a genomewide estimate of the amount of variation that can be attributed to differentiation between ethnic groups (table 4). The results presented in table 4 demonstrate that <10% of the diversity measured with this large RFLP sample can be attributed to differentiation among the ethnic groups represented in our sample.

The relative amount of allele frequency differentiation between ethnic groups was quantified by computing genetic distances among the four groups (table 5) (Nei 1971, 1972). The distances range from .066 to .096, and a phenetic UPGMA analysis coupled CA and AA samples together, separate from an AS-AI association. These associations would be predicted, since AAs have experienced gene exchange with CAs in the New World and because AIs are thought to have originated from Asian migrations 12,000–14,000 years ago (Wallace et al. 1985; Schurr et al. 1990). Further phylogenetic conclusions, particularly about topology rooting and timing of separations, would be difficult because of the limits of using only four ethnic groups, at least one of which (AA) has recently experienced gene flow.

Discussion

The present survey and previous studies of allelic variation within and between human ethnic groups indicate

that the majority of variation involves allelic frequencies and not wholesale substitutions of one allele for another (Lewontin 1972; Nei 1977; Bowcock et al. 1987, 1991). To put human molecular evolution into perspective, consider a heterozygous pair of allelic nucleotides in a random outbred individual. Coalescence theory (Hudson 1983; Tajima 1983; Hudson and Kaplan 1985) suggests that the pair of allelic nucleotides would be derived from a common ancestor $\sim 2N_e$ generations ago, where N_e is the genetically effective population size. If human $N_e = 10,000$, and generation time is 20 years, as some authors have suggested (Nei 1987; Klein et al. 1993), then the pair of nucleotides last shared a common ancestor some 400,000 years ago. If most allele polymorphisms are on the order of 400,000 years old, and if the oldest racial separation occurred 166,000–249,000 years ago (Cann et al. 1987; Vigilant et al. 1991), then we might expect a significant fraction of human polymorphisms to predate any racial separation. This expectation is borne out by our data, in that most loci polymorphic in CAs are simultaneously polymorphic, for the same pair of alleles, in the other three populations.

Population genetics theory suggests three primary forces causing the genetic divergence among populations: (1) mutational acquisition of new alleles, (2) selective differences in different environments, and (3) stochastic flux of gene frequencies that is due to genetic drift, generally under small population size. Undoubtedly, human population sizes were many orders of magnitude smaller than at pres-

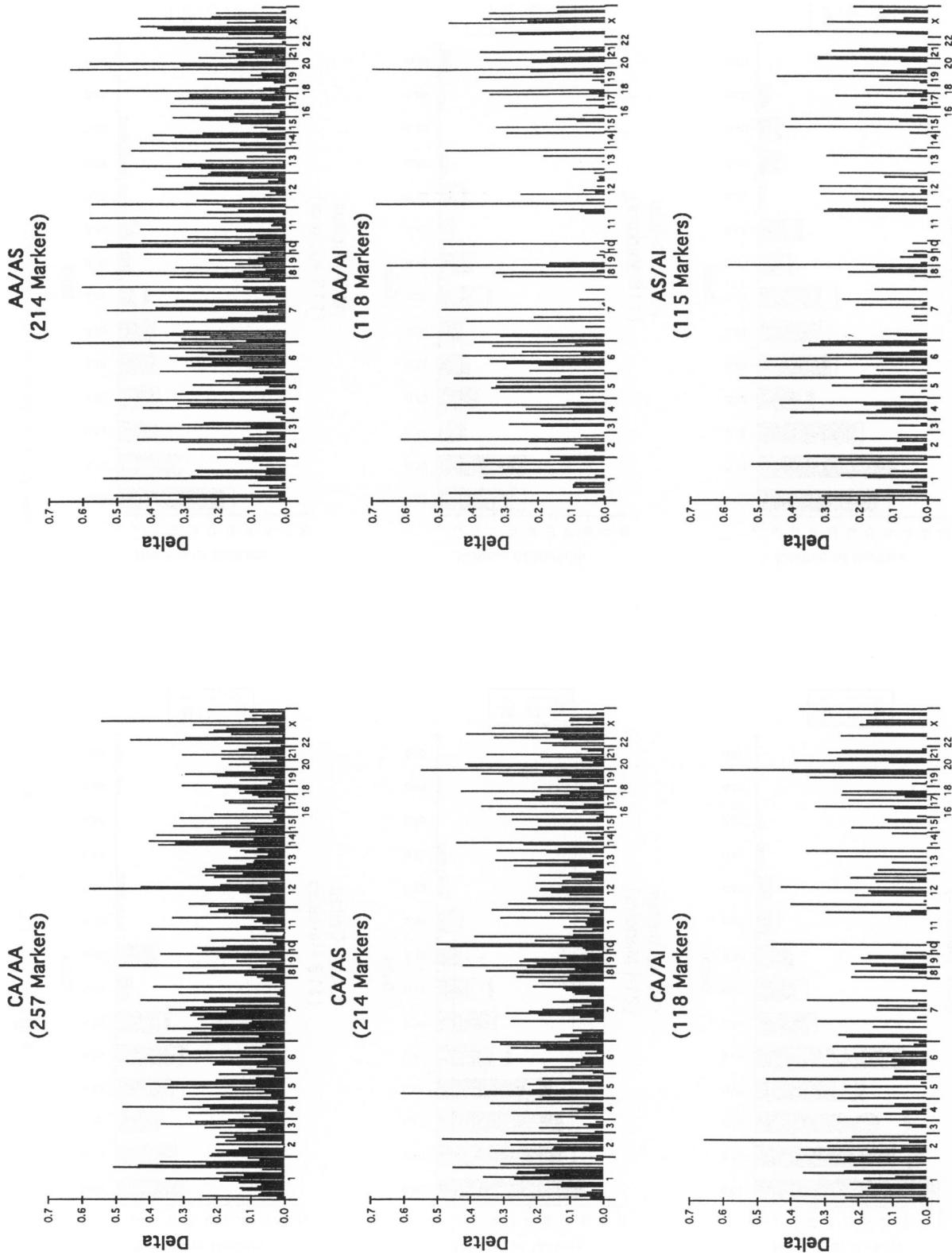


Figure 2 δ Values for all tested markers aligned according to chromosomal location on the PATM^{AP}. Chromosomal assignment is on the X-axis.

Figure 3 Distribution of δ values for all markers, between different ethnic group comparisons, from data in table 2

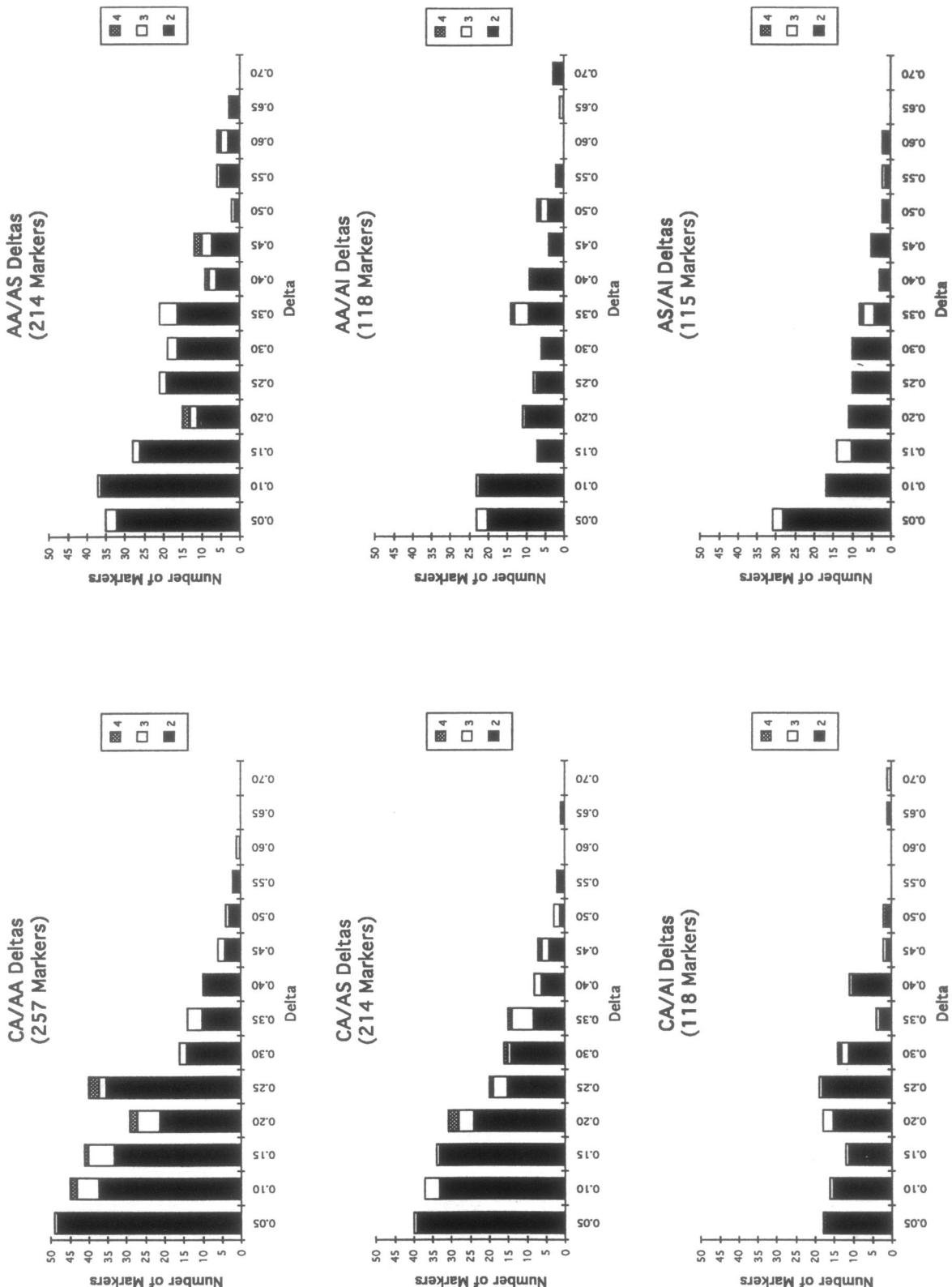


Table 3**Estimates of Average Heterozygosity h_p in Each of Four Ethnic Groups**

ETHNIC GROUP	NO. OF PATIENTS ^a	NO. OF MARKERS	h_p	
			EQUATION (1)	EQUATION (2)
CA	595	257	.399	.868
AA	206	257	.361	.799
AS	59	214	.328	.714
AI	82	118	.297	.652
Overall (h_s)	942	257	.356	.764

^a Note that not all markers were scored on each patient (see table 1).

ent, suggesting a large role for genetic drift in the currently observed differentiation. The existence of alleles that are found in some, but not all, populations may also be due to genetic drift. Alternatively, ongoing mutation may produce alleles that are restricted to a subset of the populations. Note that we have taken a conservative approach in treating RFLP alleles as identical by descent. Sequence-level variation is undoubtedly masked by this approach, which means that our δ values may underestimate the differentiation. Additionally, our selection of markers known to be heterozygous in CAs precludes finding markers with $\delta \gg .5$, if they exist. Migration and gene flow between populations will also reduce genetic differentiation, compared with that of the anthropological founders. Presumably this is the case for AAs.

In our sample of 257 markers and four races, we have seen statistically significant differences among allele frequencies at over half of the loci in any given population comparison. The magnitude of these δ values is encouraging for mapping by admixture linkage disequilibrium, which relies on large δ values for identifying genomic re-

Table 5**No. of Loci Compared (above the Diagonal) and Nei's (1972) Genetic Distances^a Among Four Human Ethnic Groups (below the Diagonal)**

	CA	AA	AS	AI
CA		257	214	118
AA068		214	118
AS066	.096		115
AI073	.098	.071	

^a Genetic distance, D , was calculated as $-\ln I$ where $I = J_{xy}/\sqrt{J_x J_y}$. I equals the normalized identity of genes between populations X and Y. D measures the average number of gene differences, per locus, between individuals from the two test populations. J_{xy} is the arithmetic mean of $j_{xy} = \sum_i x_i y_i$ over all loci; J_x is the arithmetic mean of $j_x = \sum_i x_i^2$ over all loci; and x_i (or y_i) is the frequency of the i th allele in the first (or second) population.

gions harboring genes of interest (Stephens et al. 1994 [in this issue]). Our sample of markers was chosen to be highly informative in CAs, which creates a bias if we are trying to estimate overall genomic heterozygosity or heterozygosity in non-CA groups. However, our data show that nearly all these loci are polymorphic in samples representing other racial groups and that appreciable differentiation exists, locus by locus, among these human populations. We see that in all cases the majority of measured diversity (>90%) is within, not between, these groups, as estimated with smaller gene product data sets (Lewontin 1972; Nei and Roychoudhury 1982).

Acknowledgments

We thank the many colleagues who sent us probes, as indicated in the footnotes of table 1. We are also grateful to Bruce Weir, Mike Clegg, Ray White, Mark Leppert, and Web Cavenee for discussions and suggestions. We are grateful to the investigators and participants of the multicenter AIDS Cohort Study (MACS), the Multicenter Hemophilia Cohort Study (MHCS), and the DCG AIDS Study for providing samples used in this analysis.

References

- Anderson MA, Gusella JF (1984) Use of Cyclosporin A in establishing Epstein-Barr virus-transformed human lymphoblastoid cell lines. *In Vitro* 20:856-868
- Beasley RP, Hwang LY, Lin CC, Chien CS (1981) Hepatocellular carcinoma and hepatitis B virus: a prospective study of 22,707 men in Taiwan. *Lancet* 2:1129-1133
- Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314-331
- Bowcock AM, Bucci C, Hebert JM, Kidd JR, Kidd KK, Friedlaender JS, Cavalli-Sforza LL (1987) Study of 47 DNA markers in five populations from four continents. *Gene Geogr* 1:47-64

Table 4**Apportionment of Human RFLP Genetic Variation in Four Ethnic Groups**

GROUPS COMPARED	NO. OF MARKERS	PROPORTION OF DIVERSITY, $1 - (\bar{h}_p/h_s)$, BETWEEN GROUPS, BY	
		EQUATION (1)	EQUATION (2)
CA/AA	257	.047	.044
CA/AS	214	.055	.068
CA/AI	118	.070	.080
AA/AS	214	.076	.086
AA/AI	118	.086	.091
AS/AI	115	.062	.076
CA/AA/AS	214	.079	.085
CA/AA/AS/AI	115	.095	.101

- Bowcock AM, Kidd JR, Mountain JL, Hebert JM, Carotenuto L, Kidd KK, Cavalli-Sforza LL (1991) Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. *Proc Natl Acad Sci USA* 88:839–843
- Bowcock AM, Klinger K, Bakker E, Pearson PL, Chipperfield MA, Ceverha P, Minter A (1993) Report of the DNA committee—including highly informative markers. *Genome Priority Rep* 1:810–884
- Briscoe D, Stephens JC, O'Brien SJ (1994) Linkage disequilibrium in admixed populations: applications in gene mapping. *J Hered* 85:59–63
- Cann RL, Stoneking M, Wilson AL (1987) Mitochondrial DNA and human evolution. *Nature* 34:31–36
- Cavalli-Sforza LL, Menozzi P, Piazza A (1993) Demic expansions and human evolution. *Science* 259:639–645
- Chakraborty R, Weiss KM (1988) Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proc Natl Acad Sci USA* 85:9119–9123
- Collins F, Galas D (1993) A new five-year plan for the U.S. Human Genome Project. *Science* 262:43–46
- Cuticchia AJ, Chipperfield MA, Porter CJ, Kearns W, Pearson PL (1993) Managing all those bytes: the Human Genome Project. *Science* 262:47–49
- Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 12: 241–253
- Gibbons A (1993) Geneticists trace the DNA trail of the first Americans. *Science* 259:312–313
- Ginzburg HM, Fleming PL, Miller KD (1988) Selected public health observations derived from the multicenter AIDS cohort study. *J Acquir Immune Defic Syndr* 1:2–7
- Goedert JJ, Kessler CM, Aledort LM, Biggar RJ, Andes WA, White GS, Drummond JE, et al (1989) A prospective study of human immunodeficiency virus type 1 infection and the development of AIDS in subjects with hemophilia. *N Engl J Med* 321:1141–1148
- Goedert JJ, Weiss SH, Eyster ME, Melby M, Wilson S, Ginzburg HM, Biggar RJ, et al (1986) Three-year incidence of AIDS in five cohorts of HTLV-III-infected risk group members. *Science* 231:992–995
- Hadler SC, Francis DP, Maynard JE, Thompson SE, Judson FN, Echenberg DF, Ostrow DG, et al (1986) Long-term immunogenicity and efficacy of hepatitis B vaccine in homosexual men. *N Engl J Med* 315:209–214
- Hudson RR (1983) Properties of a neutral allele model with intragenic recombination. *Theor Popul Biol* 23:183–201
- Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111:147–164
- Klein J, Takahata N, Ayala FJ (1993) MHC polymorphism and human origins. *Sci Am* 269:78–83
- Lewontin RC (1972) The apportionment of human diversity. *Evol Biol* 6:381–398
- Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet* 44:397–401
- McAlpine PJ, Shows TB, Boucheix C, Pericak-Vance MA, Anderson WA (1993) The 1992 catalog of mapped genes and report of the nomenclature committee. *Genome Priority Rep* 1: 11–142
- McKusick VA (1991) Mendelian inheritance in man, 6th ed. The Johns Hopkins University Press, Baltimore
- Murphy J, Jason J, Holman R, Gomperts E, Evatt B (1991) Immune effects of HIV on hemophiliac children in a multicenter cohort. *Proc Int Conf AIDS* 7:268
- Murray JC (1993) Microsatellite loci. *Cooperative Hum Linkage Cent* 1:1–2
- Nei M (1971) Identity of genes and genetic distance between populations. *Genetics Suppl* 68:47
- (1972) Genetic distance between populations. *Am Nat* 106:283–292
- (1977) F-statistics and analysis of gene diversity in subdivided populations. *Ann Hum Genet* 41:225–233
- (1987) Molecular evolutionary genetics. Columbia University Press, New York
- Nei M, Roychoudhury AK (1982) Genetic relationship and evolution of human races. *Evol Biol* 14:1–59
- NIH/CEPH Collaborative Mapping Group (1992) A comprehensive genetic linkage map of the human genome. *Science* 258:67–86
- (1993) A comprehensive genetic linkage map of the human genome In: O'Brien SJ (ed) *Genetic maps*, 6th ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp 5.46–5.105
- Roberts L (1992) Anthropologists climb (gingerly) on board. *Science* 258:1300–1301
- Schurr TG, Ballinger SW, Gan Y-Y, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, et al (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. *Am J Hum Genet* 46:613–623
- Stephens JC, Briscoe D, O'Brien SJ (1994) Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. *Am J Hum Genet* 55:809–824 (in this issue)
- Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437–460
- Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. *Science* 253:1503–1507
- Wallace DC, Garrison K, Knowler WC (1985) Dramatic founder effects in Amerindian mitochondrial DNAs. *Am J Phys Anthropol* 68:149–155
- Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* 44:388–396
- Weber J, Wong C (1993) Mutation of human short tandem repeats. *Hum Mol Genet* 2:1123–1128
- Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, et al (1992) A second-generation linkage map of the human genome. *Nature* 359:794–801